

METHODOLOGY

Open Access

The impact of sample storage time on estimates of association in biomarker discovery studies

Karl G Kugler¹, Werner O Hackl¹, Laurin AJ Mueller¹, Heidi Fiegl², Armin Graber^{1,3}, Ruth M Pfeiffer^{4*}

Abstract

Background: Using serum, plasma or tumor tissue specimens from biobanks for biomarker discovery studies is attractive as samples are often readily available. However, storage over longer periods of time can alter concentrations of proteins in those specimens. We therefore assessed the bias in estimates of association from case-control studies conducted using banked specimens when marker levels changed over time for single markers and also for multiple correlated markers in simulations. Data from a small laboratory experiment using serum samples guided the choices of simulation parameters for various functions of changes of biomarkers over time.

Results: In the laboratory experiment levels of two serum markers measured at sample collection and again in the same samples after approximately ten years in storage increased by 15%. For a 15% increase in marker levels over ten years, odds ratios (ORs) of association were significantly underestimated, with a relative bias of -10%, while for a 15% decrease in marker levels over time ORs were too high, with a relative bias of 20%.

Conclusion: Biases in estimates of parameters of association need to be considered in sample size calculations for studies to replicate markers identified in exploratory analyses.

Background

Using specimens, including serum, plasma or tumor tissue, from biobanks is attractive for biomarker studies, as samples are readily available. For example, archived patient tissue specimens from prospective clinical trials can be used for establishing the medical utility of prognostic or predictive biomarkers in oncology [1]. Convenience samples from clinical centers and hospitals may be of use in biomarker discovery studies. The National Cancer Institute maintains a website <http://resresources.nci.nih.gov> that lists human specimen resources available to researchers, including specimens and data from patients with HIV-related malignancies, a repository of thyroid cancer specimens and clinical data from patients affected by the Chernobyl accident, normal and cancerous human tissue from the Cooperative Human Tissue Network (CHTN) and blood samples to validate blood-based biomarkers for early diagnosis of lung cancer. However, freezing specimens over long periods of time can alter levels of some of their components [2] causing

decreases or increases in marker concentrations [3-5]. Among other factors, storage temperature [6-8] and storage time [3,9,10] are known to impact frozen samples. Thus, even in carefully collected and stored samples time alone can alter marker levels.

Our work was motivated by a biomarker discovery study at the Medical University of Innsbruck that aims to identify biomarkers to predict breast cancer recurrence. In that study, among other investigations frozen serum samples from women diagnosed with breast cancer at the Medical University of Innsbruck Hospital between 1994 and 2010 will be used to identify candidate markers that predict breast cancer recurrence within five years of initial diagnosis. These markers will then be validated in prospectively collected specimens.

While the focus of discovery is the testing of association of markers with outcome, sample size considerations for validation studies are often based on estimated effect sizes seen in discovery studies. Any substantial bias in the effect sizes seen in the discovery effort will thus result in sample sizes of the follow up study that are too small (if associations are overestimated) or lead to the analysis of too many costly biospecimens (if estimates are too low). Additionally, degradation in markers

* Correspondence: pfeiffer@mail.nih.gov

⁴Biostatistics Branch, Division of Cancer Epidemiology and Genetics, National Cancer Institute, Bethesda, MD 20892, USA
Full list of author information is available at the end of the article

could lead to missed associations, i.e. increased numbers of false negative findings, as effects may be attenuated.

We used simulations to systematically assess the impact of changes in marker levels due to storage time on estimates of association of marker levels with outcome in case-control studies. Our simulations are based on parameters obtained from data from a small laboratory experiment, designed to assess the impact of degradation on measurements of two serum markers. We study two set-ups for our simulations, one when single markers are analyzed, and one situation when multiple markers are used. While the choices of parameters depend on the specific setting, our results can help to assess the potential magnitude of a bias in and to interpret findings from studies that use biospecimens stored over long periods of time.

Methods

Markers

Cancer antigen 15-3 (CA 15-3) is a circulating tumor marker which has been evaluated for use as a predictive parameter in breast cancer patients indicating recurrence and therapy response. CA 15-3, the product of *MUC1* gene, is aberrantly over expressed in many adenocarcinomas in an underglycosylated form and then shed into the circulation [11]. High concentrations of CA 15-3 are associated with a high tumor load and therefore with poor prognosis [12]. Thus, post-operative measurement of CA 15-3 is widely used for clinical surveillance in patients with no evidence of disease and to monitor therapy in patients with advanced disease. Cancer antigen 125 (CA125), another mucin glycoprotein, is encoded by the *MUC16* gene. Up to 80% of epithelial ovarian cancers express CA125 that is cleaved from the surface of ovarian cancer cells and shed into blood providing a useful biomarker for monitoring ovarian cancer [13].

Laboratory Methods

There are numerous reports on the impact of storage time on levels of individual components measured in serum in the literature [3,5,8,10,14,15]. We selected two well-known markers and measured their degradation over time. CA 15-3 and CA-125 were determined using a microparticle enzyme immunoassay and the Abbott IMx analyzer according to the manufacturers' instructions. Serum samples were collected at the Medical University of Innsbruck, Austria, between 1997 and 2001. Sample analysis was performed first at sample collection (1997 - 2001) and then again in September 2009, after storage at -30°C until 2004 and at -50°C thereafter. Eleven samples were analyzed for CA 15-3, and nine for CA125. Of the nine samples three had CA125 measurements below the detection limit of the assay. These

samples were not used when computing mean and median differences.

Table 1 shows the values of the markers measured at the time of collection and the corresponding values for the same samples measured in September 2009.

Statistical Model

Single Marker Model

Let Y_i be one if individual i experiences the outcome of interest, i.e. is a case, and zero otherwise and let X_i be the values of a continuous marker for person i . We assume that in the source population that gives rise to our samples, the probability of outcome is given by the logistic regression model

$$P(Y_i = 1|X_i) = \frac{\exp(\mu + \beta X_i)}{1 + \exp(\mu + \beta X_i)}. \quad (1)$$

The key parameter of interest is the log-odds ratio β that measures the increase in risk for a unit increase in marker levels.

Table 1 Marker Concentration Changes

Date of sample collection	Concentration measured		% change
	at sample collection	Sept 2009	
CA 15-3			
Nov 1997	166	187	12.65
Oct 1998	29	33	13.79
Apr 1995	10	12	20.00
Feb 2001	21	19	-9.52
Apr 2001	23	24	4.35
Feb 1999	33	34	3.03
Sep 2000	26	33	26.92
Sep 2000	24	33	37.50
Sep 2000	15	17	13.33
Sep 2000	12	16	33.33
Nov 1999	884	986	11.54
CA125			
Feb 1999	83	96	15.66
Feb 1999	< LOD [†]	< LOD	
Feb 1999	< LOD	< LOD	
Feb 1999	51	69	35.29
Feb 1999	< LOD	< LOD	
Sep 2000	77	73	-5.19
Sep 2000	33	32	-3.03
Sep 1998	106	105	-0.94
Oct 1998	1273	2026	59.15

[†]LOD = limit of detection.

Concentrations of two markers, CA 15-3 and CA125, measured at the time of freezing and then again after a long term storage. Measurements with concentrations below the limit of detection were excluded from further analysis.

We assume that the biomarkers are measured in retrospectively obtained case-control samples, as this is practically the most relevant setting. That is, first n individuals with the outcome of interest ("cases") and n individuals without that outcome ("controls") are sampled based on their outcome status, and then their corresponding marker values X are obtained. In our motivating example cases are women who experience a breast cancer recurrence within five years of initial breast cancer diagnosis, and controls are breast cancer patients without a recurrence in that time period.

Storage Effects on Marker Measurements

Instead of the true marker measurement X , we observe the value Z_t of the marker after the sample has been frozen for t time units, e.g. months or years. We assume that Z_t relates to X through the linear relationship

$$Z_t = Xb_t + \varepsilon. \quad (2)$$

The additive noise is assumed to arise from a normal distribution $\varepsilon \sim N(0, \sigma_\varepsilon^2)$. Without loss of generality we focus on discrete time points, $t = 0, 1, 2, \dots, t_{max} = 10$ in our simulations. In the laboratory experiments, the marker levels for CA 15-3 increased by about 15% over a period of 10 years (Table 1). Because no intermediate measurements are available from our small laboratory study, the true pattern of change over time is unknown. Thus, we used three different sets of coefficients $b_{j,t}$ with $j = 1, 2, 3$, reflecting linear, exponential and logarithmic changes for the marker levels over time. Each set of coefficients was chosen to result in an increase of 15% after ten years of storage.

For the linear function, $b_{1,t}^i$, the yearly increase in marker levels was set to 1.5%. To model the non-linear increases in marker levels, we estimated coefficients $b_{2,t}^i$ and $b_{3,t}^i$ based on an approximated Fibonacci series f_t , where $f_0 = 0, f_1 = 1, f_2 = 2$, and $f_t = f_{t-1} + f_{t-2}$ for $t = 2, \dots, 10$. For the exponential function $b_{2,t}^i$ we normalized f_t so that $f_{t_{max}}$ was 15%.

$$b_{2,t}^i = 100 + 0.15f_t \frac{100}{f_{t_{max}}}. \quad (3)$$

For a logarithmic increase we used coefficients

$$b_{3,t}^i = 100(1 + 0.15) - b_{2,t_{max}-t}^i. \quad (4)$$

To simulate decreases in marker values over time, we used $b_4^d = -b_1^i$, $b_5^d = -b_2^i$, $b_6^d = -b_3^i$. All of these functions are plotted in Figure 1.

It is also possible to analytically assess the bias in estimates of in (1) when Z_t is used instead of the true marker value X to estimate the association with disease.

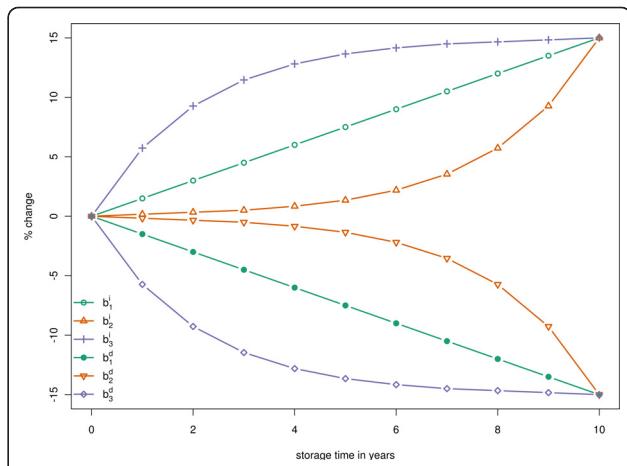


Figure 1 Choices of b_t . Three functions b_t^i model an increase in marker levels of 15% at $t = 10$, and three function b_t^d model a decrease of 15% at $t = 10$.

From (2) we get that X conditional on the measured Z_t has a normal distribution, $X|Z_t \sim N(Z_t/b_t, \sigma_{X|Z}^2)$, where $\sigma_{X|Z}^2 = \sigma_\varepsilon^2/b_t^2$. Then using results from Carroll et al. [16]:

$$\text{logit}(P(Y = 1|Z_t)) \approx \frac{\mu + \beta/b_t Z_t}{(1 + \beta^2 \sigma_{X|Z}^2/1.7)^{1/2}}, \quad (5)$$

Where $\text{logit}(x) = \ln\{x/(1-x)\}$. For multiple, correlated markers, which we study in the next section, a closed form analytical expression equivalent to (5) is not readily available.

Multiple Markers Model

We also studied a practically more relevant setting, namely that multiple markers are assessed in relation to outcome. We generated samples of $p = 10$ markers $X = (X_1, \dots, X_p)$ from a multivariate normal distribution, $X \sim MVN(0, \Omega)$. We studied two choices of covariance structure: first, we let $\Omega = (\omega_{ij})$ be the identity matrix, and second we assumed that the markers were equally correlated, with $\text{corr}(X_i, X_j) = \rho$, $i \neq j$ for various choices of ρ .

We first assumed that only one marker, X_1 , was truly associated with outcome Y , and simulated Y from the model

$$\text{logit } P(Y = 1|X_1) = \mu + \beta X_1. \quad (6)$$

We also then let three of the markers, X_1, X_2 and X_3 , be associated with the outcome,

$$\text{logit } P(Y = 1|X_1, X_2, X_3) = \mu + \sum_{i=1}^3 \beta_i X_i. \quad (7)$$

In the simulations we let each marker change over time based on equation (2) independently of the other markers for $t = 0, 1, 2, \dots, t_{max} = 10$. For X_1 the change

over ten years was 15%, and for each of the other markers we randomly selected a coefficient b_{it} from a uniform distribution on the interval [-0.2, 0.2] and used the chosen b_{it} in equation (2). We thus allowed only increases or decreases of 20% or less over ten years.

Simulations

To obtain case-control samples, we first prospectively generated a cohort of markers and outcome values (Y_i, X_i), $i = 1, \dots, N$. We drew X_i from a normal distribution, $X \sim N(0, 1)$, and then generated Y_i given X_i from a binomial distribution with $P(Y_i = 1|X_i)$ given in equation (1) for $i = 1, \dots, N$. We then randomly sampled n cases and n controls from the cohort to create our case-control sample.

For the single marker setting, we then fit a logistic regression model with Z_t instead of X to the case-control data,

$$\text{logit } P(Y = 1|Z_t) = \mu_t + \beta_t^* Z_t, \quad (8)$$

and obtained the maximum likelihood estimate (MLE) $\hat{\beta}_t^*$ that characterizes the association of outcome with the marker measured after time t in storage.

For each setting of the parameters and for each choice of b_t in (2), we simulated 1000 datasets for each sample size, $n = 75$ and $n = 200$ cases and the same number of controls for the single marker simulations, and $n = 250$ and $n = 500$ for the multiple marker settings. We also fit a logistic regression model based on the marker level X at time $t = 0$ that corresponds to no time related change in marker levels.

For the multiple marker setting, we analyzed the data using two different models. First, we fit separate logistic regression models for each marker,

$$\text{logit } P(Y = 1|Z_{k,t}) = \mu_t + \beta_{k,t}^* Z_{k,t}, \quad k = 1, \dots, p \quad (9)$$

We also estimated regression coefficients for every time step from a joint model,

$$\text{logit } P(Y = 1|Z_1, Z_2, \dots, Z_p) = \mu_t + \sum_{k=1}^p \beta_{k,t}^* Z_{k,t}. \quad (10)$$

In addition to the bias, we also assessed the power to identify true associations. When we fit separate models (9), we used a Bonferroni corrected type 1 error level $\alpha = 0.05/p$ to account for multiple testing. For the setting (10) we tested the null hypothesis $H_0 : \beta_1^* = \dots = \beta_p^* = 0$ using a chi-square test with p degrees of freedom. Letting $\hat{\beta} = (\hat{\beta}_1, \dots, \hat{\beta}_p)$ be the vector of parameter estimates of the coefficients in (10), and $\hat{\Sigma}$ denote the corresponding estimated covariance matrix, we computed

$$T = \hat{\beta}^* \hat{\Sigma}^{-1} \hat{\beta}^* \sim \chi_p^2. \quad (11)$$

Of course model (10) can only be fit to data when p is substantially smaller than the available sample size, while model (9) does not have this limitation. For the multivariate simulations we computed the power, that is the number of times the null hypothesis is rejected over all simulations.

Results

Laboratory Experiment

On average both CA 15-3 and CA125 levels increased with increasing time in storage, CA 15-3 levels increased by 15.18% (standard error 4.14) and CA125 16.82% (standard error 10.533) over approximately ten years (Table 1). This increase is most likely due to evaporation of sample material attributed to the usage of sample tubes with tops that did not seal as well as the newer ones. A similar evaporating effect was reported by Burtis et al. [17]. Alternatively, the standard used for the calibration of the assay may have decreased over the years, resulting in higher levels for the more recent analysis.

Simulation Results

Single Marker Results

We simulated storage effects for a period of ten years for three functions (b_1^i, b_2^i, b_3^i) that resulted in a 15% increase of marker levels after $t = 10$ years, and three functions, (b_1^d, b_2^d, b_3^d) , that resulted in 15% decrease after $t = 10$ years. We let $\mu = -3$ and $\beta = 0.3$ in model (1) that describes the relationship between the true marker levels and outcome. The error variance in model (2) for the change of the marker over time was $\sigma_e^2 = 0.01$. We analyzed the simulated data at three time points, at sample collection ($t = 0$), and after $t = 5$ and $t = 10$ years.

Table 2 shows the results for functions b_1^i, b_2^i, b_3^i , that result in increases of marker levels and b_1^d, b_2^d, b_3^d , that cause decreases of marker levels. The results in Table 2 are means over 1, 000 repetitions for each choice of sample size. Table 2 also shows the relative bias, computed as $\text{rel.bias} = (\beta - \hat{\beta}_t^*)/\beta$. As expected, the true association parameter $\beta = 0.3$ in (1) was estimated without bias for $t = 0$ for all sample sizes. For $t = 5$, the relative bias ranged from 2% for b_2^i to -9% for b_3^i for $n = 75$ cases and controls, and from 1% for b_2^i to -10% for b_3^i for $n = 200$ cases and controls. The small positive bias for $t = 5$ for b_2^i was not seen when the simulation was repeated with a different seed. The differences in relative bias reflect the differences in the shape of increase of marker values. As all functions were chosen to cause a

Table 2 Univariate Marker Results

<i>n</i> = 75						<i>n</i> = 200					
increase over time			decrease over time			increase over time			decrease over time		
t = 0											
<i>b</i> ₁ ⁱ	<i>b</i> ₂ ⁱ	<i>b</i> ₃ ⁱ	<i>b</i> ₁ ^d	<i>b</i> ₂ ^d	<i>b</i> ₃ ^d	<i>b</i> ₁ ⁱ	<i>b</i> ₂ ⁱ	<i>b</i> ₃ ⁱ	<i>b</i> ₁ ^d	<i>b</i> ₂ ^d	<i>b</i> ₃ ^d
$\hat{\beta}_0$	0.309	0.309	0.309	0.309	0.308	0.308	0.308	0.308	0.307	0.307	0.308
se.emp	0.005	0.005	0.005	0.005	0.005	0.005	0.003	0.003	0.003	0.003	0.003
rel.bias	0.029	0.029	0.029	0.03	0.028	0.028	0.026	0.026	0.024	0.024	0.026
rel.bias.sd	0.566	0.566	0.568	0.571	0.568	0.563	0.343	0.342	0.343	0.341	0.342
t = 5											
<i>b</i> ₁ ⁱ	<i>b</i> ₂ ⁱ	<i>b</i> ₃ ⁱ	<i>b</i> ₁ ^d	<i>b</i> ₂ ^d	<i>b</i> ₃ ^d	<i>b</i> ₁ ⁱ	<i>b</i> ₂ ⁱ	<i>b</i> ₃ ⁱ	<i>b</i> ₁ ^d	<i>b</i> ₂ ^d	<i>b</i> ₃ ^d
$\hat{\beta}_5$	0.288	0.305	0.272	0.334	0.312	0.356	0.287	0.304	0.271	0.331	0.312
se.emp	0.005	0.005	0.005	0.006	0.005	0.006	0.003	0.003	0.003	0.003	0.004
rel.bias	-0.041	0.015	-0.092	0.112	0.042	0.186	-0.044	0.013	-0.096	0.105	0.039
rel.bias.sd	0.527	0.559	0.5	0.617	0.576	0.65	0.319	0.337	0.302	0.368	0.346
t = 10											
<i>b</i> ₁ ⁱ	<i>b</i> ₂ ⁱ	<i>b</i> ₃ ⁱ	<i>b</i> ₁ ^d	<i>b</i> ₂ ^d	<i>b</i> ₃ ^d	<i>b</i> ₁ ⁱ	<i>b</i> ₂ ⁱ	<i>b</i> ₃ ⁱ	<i>b</i> ₁ ^d	<i>b</i> ₂ ^d	<i>b</i> ₃ ^d
$\hat{\beta}_{10}$	0.269	0.269	0.269	0.362	0.361	0.361	0.268	0.268	0.268	0.36	0.361
se.emp	0.005	0.005	0.005	0.006	0.006	0.006	0.003	0.003	0.003	0.004	0.004
rel.bias	-0.103	-0.103	-0.103	0.208	0.204	0.204	-0.106	-0.106	-0.107	0.199	0.202
rel.bias.sd	0.493	0.493	0.495	0.671	0.667	0.66	0.298	0.297	0.298	0.4	0.401

Mean values of the maximum likelihood estimates $\hat{\beta}_t^*$ of $\beta = 0.3$ after $t = 0, 5$, and 10 years for the various degradation functions, with empirical (se.emp) standard error and the relative bias $\hat{\beta}^*$. Simulations were performed with $\mu = -3$, and sample sizes $n = 75$ and $n = 200$. Function b_1 corresponds to a linear change, b_2 exponential change and b_3 logarithmic change in marker levels over time.

15% increase in marker levels after $t = 10$ years, all functions resulted in the same relative bias at $t = 10$, which ranged from -10% for $n = 75$ cases and controls to -11% for $n = 200$ cases and controls. For example, at $t = 10$ instead of $\beta = 0.3$ we obtained $\hat{\beta}_{10}^* = 0.269$ for $n = 75$ cases and controls and $\hat{\beta}_{10}^* = 0.268$ for $n = 200$ cases and controls, respectively. The findings for decaying markers levels were similar. Again, no bias was detected in the estimates for $t = 0$, while the relative bias ranged from 4% for b_2^d to 18% for b_3^d for $n = 200$ cases and controls. After $t = 10$ years in storage, the relative bias was around 20% for $n = 75$ and $n = 200$ cases and controls. These results agree well with what we computed from the analytical formula (5). For all settings we studied the model based standard error estimates were similar to the empirical standard error estimates and were thus not shown.

Results were similar for $\beta = 0.5$, $\beta = 1.0$, and $\beta = -0.3$, given in Additional File 1.

Multiple Marker Results

Table 3 presents results for the multiple marker simulations, when one marker was truly associated with

outcome, but the model that was fit to the data included all ten markers simultaneously (10). The results were very similar to the single marker simulations, with biases of about 10% after ten years. Correlations among markers did not affect the results. For example, the effect estimate after five years were $\hat{\beta}_5^* = 0.285$ and 0.281 for $n = 250$ and $n = 500$ for uncorrelated markers, and $\hat{\beta}_5^* = 0.282$ and 0.278 for $n = 250$ and $n = 500$ for fairly strong correlations of $\rho = 0.5$. The power to test for association using separate test with a Bonferroni adjusted α -level was adequate only for $n = 500$ cases and $n = 500$ controls.

Table 4 shows the results when three of the ten markers were associated with disease outcome. The true association parameters in equation (7) were $\beta_1 = 0.3$, $\beta_2 = 0.2$ and $\beta_3 = 0.2$. The changes in marker levels after ten years were 15%, 20% and 10% for X_1 , X_2 and X_3 , respectively. After $t = 10$ years the bias in the association estimate for marker X_1 was similar to the single marker case, and the case when only one of ten markers was associated with outcome, with $\hat{\beta}_{1,10}^* = 0.261$, with a 13% underestimate of true risk. For the other two

Table 3 Multivariate Marker Results: A Single Marker is associated with Outcome

	uncorrelated	correlated ($\rho = 0.5$)		
	$n = 250$	$n = 500$	$n = 250$	$n = 500$
t = 0				
$\hat{\beta}_0^*$	0.305	0.302	0.303	0.298
se.emp	0.091	0.064	0.128	0.093
rel.bias	0.018	0.005	0.009	-0.005
rel.bias.sd	0.304	0.213	0.426	0.309
power [†]	0.522	0.92	0.541	0.908
t = 5				
$\hat{\beta}_5^*$	0.285	0.281	0.282	0.278
se.emp	0.085	0.059	0.119	0.086
rel.bias	-0.052	-0.064	-0.058	-0.072
rel.bias.sd	0.282	0.198	0.398	0.287
power	0.527	0.926	0.546	0.908
t = 10				
$\hat{\beta}_{10}^*$	0.266	0.263	0.264	0.261
se.emp	0.08	0.055	0.112	0.08
rel.bias	-0.114	-0.124	-0.121	-0.13
rel.bias.sd	0.266	0.185	0.372	0.268
power	0.532	0.929	0.55	0.91

Results for simulations based on a multivariate setting with 10 markers, where only X_1 is associated with disease outcome with true $\beta = 0.3$, and $\mu = -3$. Levels of X_1 increases 1.5% per year. Simulations were performed with sample sizes $n = 250$ and $n = 500$. [†] The power is calculated as the number of rejected null hypotheses over all simulations.

markers the log odds ratio estimates after ten years were $\hat{\beta}_{2,10}^* = 0.169$ and $\hat{\beta}_{3,10}^* = 0.182$, corresponding to 15.5% and 9% relative bias. The power of a test for association using a ten degree of freedom chi-square test was above 90% even for a sample size of $n = 250$ cases and $n = 250$ controls.

Discussion

In this paper we quantified the impact of changes of marker concentrations in serum over time on estimates of association of marker levels with disease outcome in case-control studies. We studied several monotone functions (linear, exponential, logarithmic) of changes over time that captured increases as well as decreases in marker levels. All functions were designed so that after ten years the change in levels was a decrease or increase by 15%. This percent change was chosen based on observations from a small pilot study. Thus, for all different functions that were used to model markers changes the bias seen in the association parameter after ten years was the same, but for intermediate time points the magnitudes of biases differed, as the amount of change varied for different functions. For a 15% increase in marker

Table 4 Multivariate Marker Results: Three Markers are associated with Outcome

	$X1$	$X2$	$X3$
true β	0.3	0.2	0.2
perc.change	0.150	0.20	0.10
b^i	1	2	3
t = 0			
$\hat{\beta}_0^*$	0.3	0.202	0.2
se.emp	0.131	0.13	0.13
rel.bias	-0.001	0.012	0.002
rel.bias.sd	0.435	0.652	0.648
power [†]		0.996	
t = 5			
$\hat{\beta}_5^*$	0.279	0.199	0.184
se.emp	0.122	0.126	0.118
rel.bias	-0.068	-0.003	-0.078
rel.bias.sd	0.405	0.630	0.591
power		0.995	
t = 10			
$\hat{\beta}_{10}^*$	0.261	0.169	0.182
se.emp	0.113	0.108	0.117
rel.bias	-0.131	-0.155	-0.090
rel.bias.sd	0.376	0.538	0.584
power		0.995	

Results for simulations based on a multivariate setting 10 with correlated markers, with 250 cases and 250 controls, $\mu = -3$, and $\rho = 0.5$. The first three markers X_1 , X_2 , and X_3 are associated with outcome. [†] The power is calculated as the number of rejected null hypotheses over all simulations. Function b_1 corresponds to a linear change, b_2 exponential change and b_3 logarithmic change in marker levels over time.

levels, estimated log-odds ratios showed a relative bias of -10%, and for a 15% decrease in marker levels, log-odds ratios were overestimated, with a relative bias of about 20%. We assessed single markers as well as multiple correlated markers. The findings were similar, regardless of correlations.

While one could avoid this problem by using fresh samples, often, in prospective cohorts serum and blood are collected at baseline and at regular time intervals thereafter, and are used subsequently to assess markers for diagnosis or to estimate disease associations in nested case-control samples. This was the design that was used by investigators participating in the evaluation of biomarkers for early detection of ovarian cancer in the Prostate, Lung Ovarian and Colorectal (PLCO) cancer screening study.

If a biased estimate of true effect sizes due to systematic changes in biomarker levels is obtained in a discovery effort, this could lead to under- or overestimation of sample size for subsequent validation studies, and thus either compromise power to detect true effect sizes, or

cause resources to be wasted. For example, for a case-controls study with one control per case to detect an odds ratio of 2.0 for a binary exposure that has prevalence 0.2 among controls with 80% power and a type one level of 5%, one needs a sample size of 172 cases and 172 controls. If the effect size is overestimated by 13%, leading to the biased odds ratio of 2.2, investigators may wrongly select 130 cases and 130 controls for the follow up study, causing the power to detect the true odds ratio of 2.0 to be 0.68.

The impact of storage effects on the loss of power to detect associations of multiple markers due to poor storage conditions was also assessed in [18], but no estimates of bias were presented in that study.

If the amount of degradation is known from previous experiments, one could attempt to correct the bias in the obtained estimates before designing follow up studies. For a small number of markers changes in concentrations over time have been reported [4,15,19]. However, such information is typically not available in discovery studies where one aims to identify novel markers. In addition, while many changes were monotonic in time [14], the number of freeze-thaw cycles [10,19,20] and changes in storage conditions can cause more drastic changes. This also happened at the Medical University of Innsbruck, where storage temperature changed from -30°C for samples stored until 2004 to -50°C for samples stored and collected after 2004.

For investigators interested in validating new markers prospectively, a small pilot study that measures levels of marker candidates identified in archived samples again in fresh samples to obtain estimates of changes in levels may help better plan a large scale effort.

We assumed that the degradation was non-differential by case-control status. However, it is conceivable that degradation in serum from cases is different than those in serum from controls. While it would be interesting to assess the impact of differential misclassification, it is difficult to obtain realistic choices for parameters that could be used in a simulation study.

In summary, our results provide investigators planning exploratory biomarker studies with data on biases due to changes in marker levels that may aid in interpreting findings and planning future validation studies.

Conclusion

The increase or decrease in markers measured in stored specimens due to changes over time can bias estimates of association between biomarkers and disease outcomes. If such biased estimates are then used as the basis for sample size computations for subsequent validation studies, this can lead to low power due to overestimated effects or wasted resources, if true effect sizes are underestimated.

Additional material

Additional file 1: Univariate Marker Results for $\beta = 0.5$, $\beta = 1$, and $\beta = -0.3$. Mean values of the maximum likelihood estimates $\hat{\beta}_t^*$ of $\beta = 0.5$, $\beta = 1$, and $\beta = -0.3$ after $t = 0, 5$, and 10 years for the various degradation functions, with empirical (*se.emp*) standard error and the relative bias of $\hat{\beta}^*$. Simulations were performed with $\mu = -3$, and sample sizes $n = 75$ and $n = 200$. Function b_1 corresponds to a linear change, b_2 exponential change and b_3 logarithmic change in marker levels over time.

Acknowledgements

This work was supported by the COMET Center ONCOTYROL and funded by the Federal Ministry for Transport Innovation and Technology (BMVIT) and the Federal Ministry of Economics and Labour/the Federal Ministry of Economy, Family and Youth (BMWA/BMWFJ), the Tiroler Zukunftsstiftung (Tzs) and the State of Styria represented by the Styrian Business Promotion Agency (SFG). We also thank Uwe Siebert for bringing the breast cancer project to our attention, and Matthias Dehmer and the reviewers for helpful comments.

Author details

¹Institute for Bioinformatics and Translational Research, University for Health Sciences, Medical Informatics and Technology, EWZ 1, 6060, Hall in Tirol, Austria. ²Department of Obstetrics and Gynecology, Innsbruck Medical University, Anichstrasse 35, 6020, Innsbruck, Austria. ³Novartis Pharmaceuticals Corporation, Oncology Biomarkers and Imaging, One Health Plaza, East Hanover, NJ 07936, USA. ⁴Biostatistics Branch, Division of Cancer Epidemiology and Genetics, National Cancer Institute, Bethesda, MD 20892, USA.

Authors' contributions

RMP conceived the simulation studies, interpreted the data, and led the drafting and writing of the manuscript. HF conceived and executed the laboratory studies and took part in editing the manuscript. KGK, WOH, and LAJM performed the simulation studies and took part in writing the manuscript. AG initiated the study, contributed to the study design, and took part in editing the manuscript. All authors read and approved the final manuscript.

Competing interests

The authors declare that they have no competing interests.

Received: 3 January 2011 Accepted: 8 March 2011

Published: 8 March 2011

References

1. Simon RM, Paik S, Hayes DF: Use of archived specimens in evaluation of prognostic and predictive biomarkers. *J Natl Cancer Inst* 2009, **101**(21):1446-1452.
2. Peakman TC, Elliott P: The UK Biobank sample handling and storage validation studies. *Int J Epidemiol* 2008, **37**(Suppl 1):i2-16.
3. Holl K, Lundin E, Kaasila M, Grankvist K, Afanasyeva Y, Hallmans G, Lehtinen M, Pukkala E, Surcel HM, Toniolo P, Zeleniuch-Jacquotte A, Koskela P, Lukanova A: Effect of long-term storage on hormone measurements in samples from pregnant women: the experience of the Finnish Maternity Cohort. *Acta Oncol* 2008, **47**(3):406-412.
4. Hannisdal R, Ueland PM, Eussen SJPM, Svartdal A, Hustad S: Analytical recovery of folate degradation products formed in human serum and plasma at room temperature. *J Nutr* 2009, **139**(7):1415-1418.
5. Männistö T, Surcel HM, Bloigu A, Ruokonen A, Hartikainen AL, Järvelin MR, Pouta A, Vääräsmäki M, Suvanto-Luukkonen E: The effect of freezing, thawing, and short- and long-term storage on serum thyrotropin, thyroid hormones, and thyroid autoantibodies: implications for analyzing samples stored in serum banks. *Clin Chem* 2007, **53**(11):1986-1987.
6. Berrino F, Muti P, Micheli A, Bolelli G, Krogh V, Scialino R, Pisani P, Panico S, Secreto G: Serum sex hormone levels after menopause and subsequent breast cancer. *J Natl Cancer Inst* 1996, **88**(5):291-296.

7. Garde AH, Hansen AM, Kristiansen J: Evaluation, including effects of storage and repeated freezing and thawing, of a method for measurement of urinary creatinine. *Scand J Clin Lab Invest* 2003, **63**(7-8):521-524.
8. Comstock GW, Alberg AJ, Helzlsouer KJ: Reported effects of long-term freezer storage on concentrations of retinol, beta-carotene, and alpha-tocopherol in serum or plasma summarized. *Clin Chem* 1993, **39**(6):1075-1078.
9. Schrohl AS, Würtz S, Kohn E, Banks RE, Nielsen HJ, Sweep FCGJ, Brünner N: Banking of biological fluids for studies of disease-associated protein biomarkers. *Mol Cell Proteomics* 2008, **7**(10):2061-2066.
10. Gao YC, Yuan ZB, Yang YD, Lu HK: Effect of freeze-thaw cycles on serum measurements of AFP, CEA, CA125 and CA19-9. *Scand J Clin Lab Invest* 2007, **67**(7):741-747.
11. Cheung KL, Graves CR, Robertson JF: Tumour marker measurements in the diagnosis and monitoring of breast cancer. *Cancer Treat Rev* 2000, **26**(2):91-102.
12. Park BW, Oh JW, Kim JH, Park SH, Kim KS, Kim JH, Lee KS: Preoperative CA 15-3 and CEA serum levels as predictor for breast cancer outcomes. *Ann Oncol* 2008, **19**(4):675-681.
13. Bast RC, Feeney M, Lazarus H, Nadler LM, Colvin RB, Knapp RC: Reactivity of a monoclonal antibody with human ovarian carcinoma. *J Clin Invest* 1981, **68**(5):1331-1337.
14. Woodrum D, French C, Shamel LB: Stability of free prostate-specific antigen in serum samples under a variety of sample collection and sample storage conditions. *Urology* 1996, **48**(6A Suppl):33-39.
15. Gislefoss RE, Grimsrud TK, Mørkrid L: Long-term stability of serum components in the Janus Serum Bank. *Scand J Clin Lab Invest* 2008, **68**(5):402-409.
16. Carroll RJ, Ruppert D, Stefanski LA, Crainiceanu CM: *Measurement Error in Nonlinear Models: A Modern Perspective*. Second edition. Chapman & Hall/CRC Monographs on Statistics & Applied Probability; 2006.
17. Burtis CA: Sample evaporation and its impact on the operating performance of an automated selective-access analytical system. *Clin Chem* 1990, **36**(3):544-546.
18. Balasubramanian R, Müller L, Kugler K, Hackl W, Pleyer L, Dehmer M, Graber A: The impact of storage effects in biobanks on biomarker discovery in systems biology studies. *Biomarkers* 2010, **15**(8):677-683.
19. Chaigneau C, Cabioch T, Beaumont K, Betsou F: Serum biobank certification and the establishment of quality controls for biological fluids: examples of serum biomarker stability after temperature variation. *Clin Chem Lab Med* 2007, **45**(10):1390-1395.
20. Paltiel L, Rønningen KS, Meltzer HM, Baker SV, Hoppin JA: Evaluation of Freeze Thaw Cycles on stored plasma in the Biobank of the Norwegian Mother and Child Cohort Study. *Cell Preserv Technol* 2008, **6**(3):223-230.

doi:10.1186/2043-9113-1-9

Cite this article as: Kugler et al.: The impact of sample storage time on estimates of association in biomarker discovery studies. *Journal of Clinical Bioinformatics* 2011 1:9.

Submit your next manuscript to BioMed Central and take full advantage of:

- Convenient online submission
- Thorough peer review
- No space constraints or color figure charges
- Immediate publication on acceptance
- Inclusion in PubMed, CAS, Scopus and Google Scholar
- Research which is freely available for redistribution

Submit your manuscript at
www.biomedcentral.com/submit

